Draft Implementation Plan for the GEO Knowledge Hub

This document is submitted by the Secretariat to the Programme Board for discussion.

1 INTRODUCTION

The GEO-XVI Plenary (November 2019, Canberra) accepted the proposal for the GEO Secretariat to develop an implementation plan for development of the GEO Knowledge Hub (GKH), in consultation with the GEOSS Infrastructure Development Task Team (GID-TT) and the GEO Programme Board. The GKH implementation plan will then be submitted to the 51st meeting of the Executive Committee for decision.

Annex A contains the initial draft (version 1) of the GKH as developed by the Secretariat and circulated to the GID-TT on 15 January 2020. Comments from the GID-TT were submitted back to the Secretariat by 21 January 2020. The Secretariat has collated all feedback from the GID-TT into a comment log and is working on responding to them. The comment log will be circulated to the Programme Board by Wednesday 29 January 2020 for consultation.

2 REQUEST TO THE PROGRAMME BOARD

The Programme Board is invited to discuss the initial draft of the GKH at its 16th meeting, taking into consideration the comments already made by the GID-TT.

On 7 February 2020, Version 2 of the GKH implementation plan will be circulated to the GID-TT and to the Programme Board. The Programme Board is invited to provide written comments on Version 2 by 14 February 2020, which will be transmitted to the meeting of the GID-TT on 21 February 2020, in Beijing.

The final draft (Version 3) of the GKH implementation plan will then be produced for decision by the 51st Executive Committee meeting on 19-20 March 2020, taking into consideration feedback received from both the Programme Board and the GID-TT.

ANNEX A

GEO KNOWLEDGE HUB

Implementation Plan

March 2020 - March 2021

VERSION 1.0 | 15/01/2020

1. Introduction

Purpose

This implementation plan outlines the development of the GEO Knowledge Hub (GKH). The GKH is a digital repository (library) providing access to knowledge required to build applications of Earth observations (EO). The purpose of the GKH is to reveal all components of a given application using EO data, including: (a) research papers and reports describing methods and results; (b) software algorithms and cloud computing resources used for processing; (c) in situ and satellite imagery data used; and (d) results for verification. The GKH is one component of the GEOSS Infrastructure and its development is included in the GEOSS Infrastructure Development Foundational Task (FT), which forms part of the 2020-2022 GEO Work Programme (GWP).

The contents of the GKH are *linked documents that contain relevant information for EO applications that promote reproducibility, scalability, and co-design/co-production*. Examples of *documents* include an HTML file, a PDF file (report or published paper), a Jupyter Notebook, an R or python markdown file, a GitHub page, a repository entry linking to a dataset store with an assigned Digital Object Identifier (DOI), an AWS or other links to datasets, Open Geospatial Consortium (OGC) service links for data, a video (see Figure 1). We also use the term *document set* to describe a set of documents linked to the same application.

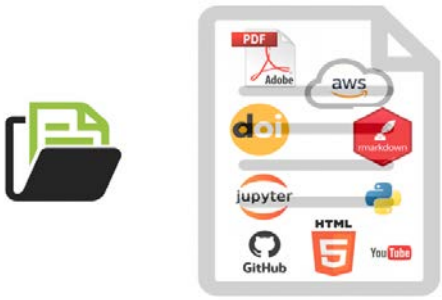

Figure 1 - Examples of documents in the GEO Knowledge Hub

Assumptions and Constraints

In the drafting of this GKH implementation plan, several general assumptions and constraints have been made that will have impacts on development of the GKH, and which need to be made explicit in the interest of transparency.

1) Vendor independence
   As defined by Wikipedia, vendor lock-in occurs when “users become dependent on a vendor for products and services, unable to use another vendor without substantial switching costs”. Thus, in order to be viewed as a transparent, neutral, honest information broker, it is essential that the GKH be independent of any particular proprietary components to offer maximum flexibility in choice. This will be

accomplished by promoting community-based, open source solutions in which all methodologies, algorithms, code/software and results are open source, and can be used in a variety of virtual machines and cloud computing platforms.

2) Reliance on pre-existing, open source, free solutions/software
The development of the GKH will benefit from the substantial ongoing effort by the open source community to develop tools for scientific repositories. A large number of universities and research institutions worldwide are now using open source software solutions which not only have the benefit of transparency and adaptability, but also greatly expedite development by avoiding duplication/reinventing the wheel while accommodating the rapid evolution of technology and societal needs.

3) Intended user
Although the GKH will be useful to a wide range of stakeholders, from national experts needing to report on policy commitments, to individual end users seeking practical solutions to local environmental challenges, the primary user of the GKH will be knowledgeable experts/technicians interested in scaling up applications of the GWP. Technical experts from research institutions may serve as intermediaries in assisting local end users to make full use of the GKH. GEO intends to leverage the capacity development networks of its partners in a "training the trainers" approach to applying the contents of the GKH.

4) Delivering knowledge
The GKH is a central digital archive providing access to codified knowledge involving the application of EO for solving societal challenges. Emphasizing a co-design/co-production strategy, the GKH is concerned with detailing all development phases of applications, elaborated in an open science environment, which support management of the environment and sustainable development. As such, *the GKH should not be viewed as delivering products or services*. Rather, the GKH represents a centralized, efficient means for transferring knowledge and scaling-up applications developed in the GWP.

5) Scaling-up applications
One of the ultimate aims of the GKH is to amplify the work of Flagships and Initiatives of the GWP such that solutions based on EO being applied at the local/national level may be scaled-up for implementation in other regions around the world, thus allowing these solutions to "go global." The GKH will focus on assembling the best in "tacit knowledge" (defined as skills, ideas and practices) gleaned from the experiences of activities of the GWP and translate it into "codified knowledge" (or knowledge that can be readily articulated, accessed and transmitted).

6) Pre-operational stage of prototype development (Potential Determined)

The GKH will serve as an intercomparison platform where different EO applications are made available and compared, including validation criteria and an evaluation of the relevant merits and demerits of each methodology and EO datasets used, for

different geographies, scales and solutions. GEO, being a global enterprise by virtue of its intergovernmental nature and convening power, is the ideal environment to test prototype development by means of the GKH whose repositories distill the broad range of knowledge gained by the GEO community over the years.

7) Resources
As a general principle, the resources (both human and financial) required to bring an open source EO application to fruition, along with its inclusion in the GKH, will fit within the existing resources of Secretariat and the contributing partners (GEO Members and/or Participating Organizations). However, this does not preclude the GEO Secretariat from seeking external sources of funding, including the commercial sector, to develop certain aspects of the GKH, and/or support application development. In these cases, any arrangements concluded with external funding sources will be impartial and made transparent to the GEO community, in accordance with *Annex C: Rules of Engagement with the Commercial Sector*, *GEO Rules of Procedure* (March 2019 update)

8) Peer review
Knowledge packages[1] under consideration for submission to the GKH must show evidence of having been peer-reviewed for reliability, either through journal publication or approval by recognized experts in the community prior to GKH publication.

9) Unrestricted access for discovery and retrieval
Access to the GKH will be unrestricted to any individual who wished so search, discover and retrieve EO knowledge packages. No account registration will be required.

10) Restricted access for publication
Since reliability and accuracy are critical characteristics of the GKH in terms of applications contained in the repositories, access to the GKH will be limited to trusted individuals only for depositing knowledge packages and making comments. In this instance, account registration (single user) will be required.

Information Flow

A simplified GKH data flow diagram is presented in Figure 1, and described in more detail in what follows.

[1] A "knowledge package" is defined here as the ensemble of EO data, methodologies, algorithms, code/software, computing environment (where relevant) and published report(s) that supports a given EO application - see Introduction.

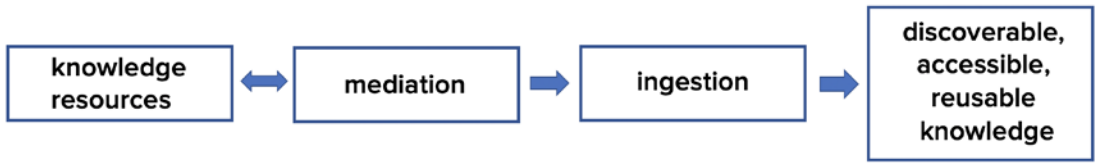

Figure 1 - Information flow in the GEO Knowledge Hub

The *knowledge resources* of the GKH include information relevant to the activities of the GWP. We expect each activity to contribute to the GKH, sharing results and best practices. This practice will be beneficial to all involved, by enhancing the visibility of GEO's work and providing a unique focal point for the results of the GWP. It will help all those interested in EO to have a place to go to learn how best to use big EO datasets. The Secretariat will work with the Programme Board to ensure that Flagships and Initiatives of the GWP contribute to expansion of the GKH.

Mediation (curation) is the process of joint work by the GEO Secretariat and the GKH contributors to ensure inclusion of trusted information. Since the GEO community needs reproducible best practices, information going into the GKH must be verified and organised, while disperse components of a document set have to be linked. The GKH team will interact with the authors so that the methods, data and software are consistent and follow the GEOSS Data Sharing principles, the GEOSS Data Management Principles and other applicable Open Science principles.

The *ingestion* process of the GKH will be as automated as possible, but human intervention is required for the final checks. For best query results, it will include the full text of documents; this requires advanced text-based search capabilities. Given the need of GEO to provide open global access, the GKH will store either open access papers or post-prints[2] of journal papers that are not open access.

Requirements

The requirements identified by the Expert Advisory Group for the GKH include:

R1. Support efficient text-based search.

R2. Link document sets with different components (web pages, PDFs, links to DOIs, videos, Jupyter Notebooks, R markdown, URLs to data, GitHub pages, videos, etc).

R3. Use descriptors compatible with current search engine technologies and emerging solutions.

R4. Be integrated with the GEO website and GEOSS Platform and based on open source software.

[2] Most scholarly publishers allow researchers to share *post-prints* in public repositories. Post-prints have the same content as the journal paper minus the formatting. A detailed list of publishers' policies is available at `http://www.sherpa.ac.uk/romeo`.

R5. Include revision services for data entry, based on contributions from the GEO community.

R6. Describe big EO data catalogues.

R7. Promote abstract description of methods used for cloud computing to facilitate interoperability.

R8. Support communities of practice to build packages that use open source scripting languages for big EO data analysis in the cloud to promote application portability.

R9. Include applications that use models and support the open sharing of modelling software.

R10. Work with the GEOSS Platform and other dataset search engines such that they access and promote repositories of continuous in situ data collection.

R11. Ensure in situ data sets stored in accredited and recommended repositories are indexed.

R12. Build an in situ data repository, managed by the Secretariat, to ensure long-term curation/mediation and preservation of data entrusted to GEO by its community.

R13. Provide links from the papers stored in the GKH to the accredited long-term data repositories.

R14. Build a data repository for long-term archival, where needed, of Citizen Science data associated with in situ observations; otherwise link with existing Citizen Science repositories for seamless retrieval of data.

R15. Promote and disseminate open source software for building multi-satellite analysis ready data that work on EO cloud platforms.

This implementation plan will ensure that each of these requirements is accounted for in the full deployment of the GKH.

In summary, the GKH will function as a digital library containing “recipe books” for reproducing and scaling up the EO applications for societal benefit globally. These open source solutions will be vendor independent and take advantage of the zero download model of the new digital economy, where big EO data is queried and analysed on the cloud, eliminating the need for costly and quickly-outdated computing infrastructures at the national level. Advances in technology, coupled with the convenience of cloud computing, thus will lower the barrier to for governments, particularly those of developing countries, to access and apply EO. In so doing, the GKH will fill the role of a robust, trusted, authoritative information for decision making.

2. Management Overview

The development of the GKH is part of the the GEOSS Infrastructure Development FT, outlined in the 2020-2022 GEO Work Programme (GWP):

> "*The Knowledge Hub is envisioned as a set of curated and linked documents that contain relevant information for Earth observation applications and which is integrated with the GEO website. It is intended to provide authoritative, validated and reproducible content for evidence-based reporting on policy commitments and decision-making*."

Also as envisaged in the GWP, the development of the GKH will be overseen by the GEOSS Infrastructure Development Task Team, whose duties are to *"ensure coordination, and integration as appropriate, among the various components of the GEOSS infrastructure for management and sharing of data, information and knowledge resources*". Among the duties of this Task Team are: "*establish performance expectations for each component and means to measure their achievement, and monitor progress of each component at regular intervals against the performance expectations*". Also, as befits all FTs of the GWP, the development of the GKH will be reported periodically to the GEO Programme Board.

Therefore, the governance of the GKH implementation has been well established as part of the GEO structures. In particular, the current document provides subsidies to the GEOSS Infrastructure Development Task Team and to the GEO Programme Board concerning the relationship of the GKH and the other components of the GEOSS infrastructure. More specifically, the performance expectations of the GKH and the progress monitoring milestones are discussed in Section 6 of this document.

Considering the actual implementation of GKH, the current document assumes a minimal team that does not require additional resources than those listed in the "GEO Secretariat Concept of Operations" (CONOPS), whose tasks are outlined in the Table below (please refer to the CONOPS document for more details).

ROLE	NAME	Roles
GEO Secretariat Director	Gilberto Camara	1. Continuously assess the status and progress of the GKH 2. Connect with Member States and Participating Organizations to identify ways that GKH can support their needs 3. Evaluate how well GKH is meeting its goals and objectives.
Senior Scientist	Douglas Cripe	4. Has primary responsibility for the Secretariat contribution to GEOSS Implementation FT. 5. Lead in planning and coordination of the GKH implementation. 6. Identify projects of the GWP to most likely to produce results relevant to the GKH.

ROLE	NAME	Roles
GKH Manager	Gilberto Queiroz (seconded from Brazil)	1. Lead the Secretariat team responsible for the implementation of the GKH. 2. Support the GWP to develop the relevant contributions of each to the GKH, including reproducible research.
Data Officer	Paola De Salvo	7. Engage with activity of GWP to ensure the EO data, algorithms, models as outputs of their respective work plans are available discoverable and accessible in the GKH; 8. Contribute to the GKH by providing technical advice and guidance on content: essential EO data – algorithms and models for the various GEO Activities and GEO engagement priorities: Sendai, Climate SDGs; 9. Act as curator of GKH content in close collaboration with Content/Data providers
In-situ Data Specialist	Florian Franziskakis	1. Be responsible for the development of the in-situ data component of the GKH. 2. Engage GWP Activities, Member States and Participating Organizations to submit relevant in-situ data and associated documentation to GKH. 3. Develop software modules for ingestion, indexing, querying and presentation of in-situ data and related documents in the GKH. 4. Develop a database for in-situ data in the GKH, based on open standards. 5. Manage and curate in-situ data in GKH in accordance with GEOSS Data Management Principles.
Space Data Specialist	Vacant (duties to be fulfilled by other GEOSEC staff)	1. Be responsible for the development of the space data component of the GKH. 2. Engage GWP Activities, Member States and Participating Organizations to submit relevant information on the use of space data and associated documentation to GKH.
IT Officer	Hendrik Baeyens	1. Take part in the design and implementation of the GKH. 2. Be responsible for management and maintenance of information technology solutions used in the GKH. 3. Coordinate the use and maintenance of the GKH for relevant GEO documents.

ROLE	NAME	Roles
Capacity Development Officer	Vacant (expect secondment from Australia)	1. Support the GEO community to use the new technologies enabled by the GKH.; 2. Work with CEOS and the Open Data Cube community to support institutions in Member States to better extract information available in the GKH; 3. Organize training sessions and webinars to improve the capacity of Member States and Participant Organizations to use the new technologies of big data, cloud computing, and machine learning; 4. Support capacity development at both individual and institutional levels for long-term initiatives of the GWP (e.g., Digital Earth Africa).

3. Major Tasks

In this section, descriptions of the major components required to implement the GKH are provided. They have been identified by the GEO Secretariat, the GEOSS Infrastructure Development Task Team and the GEO Programme Board as the optimal number and design to implement the initial instance of the GKH. These tasks may be modified or supplemented as the GKH becomes established, subject to experience and feedback from the GEO community. It is anticipated that any modifications to these tasks would happen on an annual basis, in consultation with the above-mentioned bodies and subject to approval by the GEO Executive Committee. Thus, the tasks cited herein constitute the body of work to be accomplished for the period March 2020 - March 2021, with an update containing any potential course-corrections for the period March 2021 - March 2022 to be submitted to the 1st quarter Executive Committee meeting in 2021 for consideration, and so on.

Task 1: Implement GEO Knowledge Hub (GKH) Digital Library

Objectives

- Ensure that the selected GWP activities have access to the first version of the digital library (GKH) to share and discover EO applications (see tasks 3 and 4 below).

Activities

- This task will focus on customizing an existing open source digital repository (Zenodo) to provide a stable platform where knowledge providers from selected GWP activities will be able to share their EO applications (task 3) and where users will have access to them.

- Customization includes changes to the web interface, deployment of the platform in a cloud-based environment (with possibility for portability and adaptation of the computational infrastructure when needed).

- During this stage, feedback from the knowledge providers and end users will be crucial to identify improvements to the user interface and the functionalities of the initial version of the digital library. All code related to the initial version of the digital library will be hosted on Github. In parallel, user documentation will be developed to describe the knowledge ingestion process, as well as the search features and a users' forum will be established.

Human / financial Resources

This task will be conducted by the GEO Secretariat GKH Team. In order to host and deploy the GKH in a cloud-based environment, a proposal was submitted by the GEO Secretariat to AWS. This proposal was accepted and has already been used to run the proof of concept of the GKH during the Plenary. No additional or external resources are required in order to complete this task.

Key Persons

- Rik Baeyens (IT Officer, GEO Secretariat)
- Florian Franziskakis (In situ Data Specialist, GEO Secretariat)
- Gilberto Ribeiro de Queiroz (Seconded expert, Brazil)

Criteria for completion

- Task 1 will be considered as completed when the digital library is in a first and stable version used to support tasks 3 & 4. This means that the GKH website is operational, allowing providers to register knowledge packages, store resources and assign DOIs to them if needed, link to external resources such as code repositories.
- Completion of this task is also subject to users being able to search and retrieve these knowledge packages and resources.

Duration

- Timeframe for completion of the task is expected to be 2-3 months after approval of the Implementation Plan by the 51st meeting of the GEO Executive Committee (March 2020).

Task 2: Enhance GEOSS Platform Interface to the GKH

Objectives

The aim of this task is to augment the GEOSS Platform and the GEOSS Portal with the information available in the GKH. Its motivation is to create synergies between the GEOSS Platform and the GKH. It fits with the aims expressed by the GEOSS Platform team to expand the kinds of information retrievable by the GEOSS Portal interface. Besides providing access to descriptions and location of EO data, the long term aim of the GEOSS Platform is

to enable its users to have access to methods, algorithms, software and reports available in the GKH. This capability was demonstrated at the GEO Week 2019.

The general idea for this path of evolution of the GEOSS Platform is the concept of *knowledge package*. A knowledge package is a consistent set of data, software and reports that supports an EO application. Organising a knowledge package requires combining the access to data with the access to the documents required to make use of the data. It also requires that the resulting knowledge package be organised inside the GEOSS Platform so as to make it easily accessible.

On the basis of the current discussions between the GEO Secretariat and the team responsible for developing the GEOSS Platform, this task will be done under the responsibility of the GEOSS Platform team. Its development will use the API (application programming interface) of the ZENODO platform to extract information from the GKH.

Activities

The activities in this task will be further developed by the GEOSS Platform team and will be included in future versions of this document.

Human/financial Resources

The development of the GEOSS Platform has been supported by the European Commission and the European Space Agency. The team includes specialists from ESA, CNR (Italy) and JRC. The financing of the activities under this Task will be subject to discussions between the European Commission, ESA, and the GEO Secretariat.

Key Persons

- Joost van Bemmelen (ESA)
- Guido Colangeli (ESA)
- Paolo Mazetti (CNR)

Criteria for completion

- The criteria for completion of this task will be proposed by the GEOSS Platform team.

Task 3: Identifying Reproducible Applications from GEO Work Programme (GWP)

Objectives

A key feature of the GKH is that the EO applications featured therein will be conducted in an open-science context, adhering specifically to the principle of reproducibility. By reproducibility, we mean that an interested expert:

"...can re-create the final reported results of the project, including key quantitative findings, tables, and figures, given only a set of files and written instructions[3]."

Working closely with Task 4 (see below), this Task will promote and identify open science solutions with the "ingredients" and "recipes" needed to provide clear, step-by-step instructions for reproducing a given application of EO, across the full information chain (from data, methodologies and algorithms, to computing environments and results). Part of this effort will include support for the GEO community to learn how to apply the applications in a "train the trainers" approach, for activities of the GWP that wish to progress from a local/regional scale up to a global one.

Activities

1. Open-science reproducible EO applications developed by activities of the GWP will serve as the basis for populating the GKH. The GWP Flagships and Initiatives in particular are the main vehicles by which new knowledge is being developed in the GEO community. Thus, the primary work of the GEO Secretariat will be to motivate the different GWP communities to contribute to the GKH with open source reports, data and software. To this end, the GEO Secretariat will work with the leadership of the GWP Flagships and Initiatives to establish a GKH point of contact who can serve as a Liaison Officer and keep the Secretariat informed of each phase of application development within that activity. The Secretariat will also ensure that the methodologies used in application development belonging candidate knowledge packages for inclusion in the GKH have been properly vetted through a peer-review process.

2. The GEO Secretariat will also work with the commercial sector to provide opportunities to the GEO community to leverage new technologies for data analytics and cloud computing that can contribute reproducible, open-science applications of EO. Examples include the GEO-Amazon Web Services (AWS) EO Cloud Computing Programme, and the GEO-Google Earth Enging (GEE) Licencing Programme.

3. In addition to gleaning open, reproducibly EO applications from the GWP, the GEO Secretariat will actively promote the use of open science solutions based on EO, and by extension, good use of the GKH, in a "train the trainers" approach to capacity development. Activities will include electronic fora to exchange best practices, webinars, presentations, and side-events in the context of regular GEO meetings, such as the GEO Symposium and Plenary.

[3] Kitzes, J., D. Turek, and F. Deniz, 2018: *The Practice of Reproducible Research: case-studies and lessons from the data-intensive sciences.* University of California Press.

Human/Financial Resources

The contributions of the GEO Members and Participating Organizations to the GWP will form the major component of the human and financial resourcing for this task. Additional resources allocated to this task will include staff from the GEO Secretariat.

Other sources of financial resources include commercial sector contributions, such as the GEO-AWS EO Cloud Credits Programme and the GEO-GEE License Programme. Additional external funding arrangements are under discussion at the time of this writing.

Key Persons

- Douglas Cripe (Senior Scientist, GEO Secretariat)

Duration

- Ongoing.

Criteria for completion

- A systematic procedure for interaction with the Flagships and Initiatives of the GWP has been established.
- A number of robust “cookbooks” using a variety of EO and computing platforms have been identified and used by the GEO community.
- A community for best practice exchange has been established (including commercial sector projects).

Task 4: Showcasing GEO Community Resources (pre-operational) in the GKH

Objectives

The technical aspects of implementing and scaling up of tools/applications developed by activities of the GWP are dealt with. In order to populate the GKH, the GKH team will work in very close collaboration with the GWP activities identified in Task 3 (see above) to make their knowledge and tools developed are discoverable and accessible to the global community in the GKH.

Activities

The scenario we envisage is as follows considering the example of GEOGLAM Flagship:

The Sen2-Agri open source software has been developed, in the framework of GEOGLAM Flagship, which integrates data from Sentinel-2 and Landsat 8 satellites to operationally produce indices of crop development, from local to global extents.

Through the open-science environment of the GKH, the user will have access to the data, reports, methodologies and software produced by the Sen2-Agri development team on best practices for crop monitoring using big EO data.

1) All the various Knowledge components will go under a careful quality control process of the knowledge Provider and the GKH team to secure its effective discoverability, accessibility and usability.
2) The GKH team will work with the developers to make such tools available globally, either: (a) as a virtual machine that could run on cloud services such as AWS; (b) as reproducible scripts that could run on services such as GEE.
3) The knowledge and tools made available will be complemented by ancillary essential documentation such training material, publication, case studies. The scenario above described will be applicable to knowledge elements and packages developed by the various GEO WP Activities.
4) The GKH will also support Digital Earth Africa allowing the sharing of reproducible tools/methods documented and accessible developed in the framework of Digital Earth Africa.
5) A process of capacity development towards the knowledge providers in using and populating the GKH will also be part of this task, using dedicated virtual webinars as well face to face meeting.

Human/Financial Resources

The contributions of the GEO Members and Participating Organizations to the GWP will form the major component of the human and financial resourcing for this task. Additional resources allocated to this task will include staff from the GEO Secretariat.

Key Persons

- Paola De Salvo (Data Officer, GEO Secretariat)

Duration

- Ongoing.

Criteria for completion

- Successful deployment of select EO applications from the GKH available via cloud services, verified by at least one external party.

Task 5: Improvements to the KH Digital Library

Objectives

Provide an open-source digital library designed to serve the community and to ensure that all components of EO applications are visible, accessible, and where GWP Activities can share their knowledge and experience in the long term.

Activities

Based on feedback from the community (through tasks 1, 3 and 4), this task will be focused on creating a fully customized digital library that fits the GEO community needs to go towards reproducible science. This extensive customization will be a collaborative effort with the CERN development team, within the frame of their new project named InvenioRDM, a customizable research data management (RDM) framework.

The digital library will include existing basic functionalities from task 1 as well as new ones such as quality control mechanism of the resources, full-text indexing and search of stored text documents, AI and ML to interpret documents and extract metadata elements.

One of the main additions is the establishment of a quality control mechanism to ensure that all resources registered in the GKH are as reliable and pertinent as possible. Details of this feature remain to be discussed and explicitly highlighted by the GEOSS Infrastructure Development Task Team.

As feedback from the community is of paramount importance, an integrated users community forum will be established (based on the StackExchange model) will be built. Contribution guidelines for developers and open-source sharing of the code will be maintained, as well as documentation.

Human/Financial Resources

In continuation of task 1, resources allocated to this task will include staff from the GEO Secretariat. Initial deployment of the GKH will take place in a cloud-based environment and be designed to have maximum portability. Currently, financial resources include a grant from AWS to the GEO Secretariat.

At the end of this task, a cost analysis will give a better understanding of the appropriate funding mechanisms to put in place for the long-term sustainability of the GKH.

Key Persons

- Rik Baeyens (IT Officer, GEO Secretariat)
- Florian Franziskakis (In situ Data Specialist, GEO Secretariat)
- Gilberto Ribeiro de Queiroz (Seconded expert, Brazil)

Criteria for completion

- This task will be considered as completed when the GKH is ready to be made publicly available, giving the possibility to all providers to make contributions , upon registration and through quality control mechanisms.
- Completion of this task also includes allowing public users and EO practitioners to search through the fully customized operational web interface of the GKH and retrieve relevant resources.

Duration

- This task is expected to last until the end of the present implementation plan with regular updates (each 3 months).

4. Intellectual Property Rights, Security and Privacy

As a design principle, all of the content of the GKH will be available under an open source and open access license. The responsibility for defining the specific license to be used is of the knowledge provider. If required, the GKH implementation team will provide support to the knowledge providers in understanding the differences between the licenses.

For software, providers will be recommended to use one of the licences listed by the Open Source Initiative (https://opensource.org/licenses). There are three main flavors of open source licences: "copyleft" ones such as Gnu Public License (GPL) (which restrict the use of open source code within a non-open environment), "permissive" such as the Massachusetts Institute of Technology (MIT) license (which has no restrictions for its use) and dedicated ones such as the NASA agreement (which protects the rights of the US government). For EO applications developed in popular scripting languages such as R and python, the GPL licence is the most common choice. The GKH team will do its best efforts to provide guidance to providers if required.

For documents, the recommendation to providers is, whenever possible, adopt one of the Creative Commons licences: CC-BY (permissive, just asks for citation), CC-BY-SA (mandates that all derivative works be openly shared) , CC-BY-NC (forbids commercialisation) and CC-BY-SA-NC (combines the previous two restrictions). As in the case of software, the GKH will support users to choose which licence meets their needs. In any case (documents or software), there will be no transfer of intellectual property rights from their owners to GEO. Knowledge producers retain their rights.

When storing papers that have been published in scientific journals in the GKH, there are two possibilities. Either the papers are made available as open access or the journal has the copyright of the final version. In the former case, there is no restriction for depositing the paper in the GKH. In the latter situation, GKH will abide by the journal's intellectual property rights (IPR) policy. For the majority of scientific journals that allow it, GKH will store the "post-print" version (the final version of the paper without the journal's letterhead and formatting).

In terms of data security, there are two considerations. The first one concerns the integrity of the information deposited in the GKH. For this purpose, the GKH Implementation Team will use the security system enabled by the cloud provider to make sure the information is replicated in different geographic locations and backups are made on a regular basis. The second concerns the rights of access to modify the contents of the GKH. Such access will be restricted. The user interface that will be made available for depositing data in the GKH will not provide direct access to the GKH database. The documents will be deposited by the user in a temporary storage. They will be revised and checked by the GKH management before being included in the main database.

As regards privacy, one should distinguish between individual and collective privacy. As regards personal privacy, the GKH is not expected to hold personal information that goes beyond what is already shared in documents and research papers: author's names and institutional addresses. For in-situ data collection that involves personal data, as in the case of some citizen science initiatives, the GEO Secretariat will consults with legal experts and the WMO legal office to fully understand its responsibilities for hosting data produced by third parties, in the light of legislation such as the European General Data Protection Regulation.

One should also consider the case of collective privacy rights. For example, some Member States may be concerned that the GKH allows diffusion of methods and data that enable extraction of what such countries deem as sensitive. In this and similar issues, GEO has to refer to the appropriate international legislation, and especially to the 1967 Outer Space Treaty. In force since 1967, it has been ratified by 103 states. The treaty stipulates that all activities must, among other restrictions: be governed by International Law; provide free access to space with no interference from other states; only peaceful activities may be carried out; national governments are responsible for the activities of their enterprises. Furthermore, the UN Committee on Peaceful Uses of Outer Space (COPUOS) has repeatedly recognized the value of EO as a means of providing essential information for sustainable development. Recent reports by COPUOS (e.g, AC.105/1014 from 2012 and AC.105/1115 from 2016) recognise the work of GEO and the importance of open provision of data and results from EO. Therefore, the GKH will only host data and results that are consistent with the accepted international practices and recommendations. To meet such requirements, data, documents, software, and methods stored and managed by the GKH have to be fully open and accessible to all interested parties.

5. Implementation Support

Introduction

This section details technical requirements for the GKH digital library which will be used by providers to register their linked EO application resources and by end-users to retrieve them. Access to cloud computing instances and related information are not part of this section because it will be the responsibility of the providers to prepare the environments and up to users to choose where to run applications. (vendor independent)

Hardware

Initially, the GKH will run in a cloud-based environment with an infrastructure designed to be portable and vendor independent using Elastic Cloud-Computing (EC2) instances, Simple Storage Services (S3) buckets, Elastic Block Storage (EBS). It will be accessible through a public, fixed IP. The computational infrastructure sizing assessment is part of Task 1, the processing power, internal memory as well as the storage capacity for the GKH can be easily adapted to include the initial showcase working environments (Tasks 3 & 4).

Currently, the GEO Secretariat has been granted US$ 50,000 of cloud credits by AWS to host and run the GKH. This amount is expected to be enough to cover the execution of

Tasks 1, 3, 4 & 5. Data storage will be managed with Simple Storage Service (S3) and metadata will be managed within a PostgreSQL database.

Software

The GKH is being developed on top of Invenio, a free and open-source framework for building large-scale digital repositories. The Invenio project is actively developed and maintained by CERN and a large user community. Invenio relies on the following stable, widely-used, open-source technologies:

- PostgreSQL (database management system);
- Elasticsearch (index and searching);
- RabittMQ (message queue);
- Celery (background task worker);
- NGINX (web server);
- Kibana (Elasticsearch inspection and dashboard);
- Flower (Celery monitoring);
- Redis (cache management);

The GKH will run inside a Docker container that will be integrated inside a cloud-based instance. The source code will be made available via Github and will use Python, Javascript frameworks, HTML and CSS. The main development tasks will be managed via the Github issues.

6. Performance Monitoring

Unlike traditional information systems, whose performance is measured in terms of their throughput (e.g., MFlops, transactions per second, downtime) the assessment of the GKH depends essentially on its content. In that sense, performance of the GKH will depend not only on quality of the implementation but mostly on the extent of its contents. Therefore, we propose the following indicators be used to assess the performance of the GKH:

- Number of knowledge packages available.
- Number of papers deposited.
- Number of shared scripts and software modules.
- Number of in-situ data sets associated with applications.

As stated above, a knowledge package is a combination of data, software, documents, and methods which enables application reproducibility. It represents the "gold" standard for the contents of the GKH. However, not all contributions to the GKH will consist of knowledge packages. The inclusion of individual papers and of software modules is also relevant. For many contributors, it will be simpler to share documents describing an application than it is to prepare a fully organized and documented knowledge package. Therefore, the GKH should be open to all relevant contributions.

Considering the current state of the GEO Work Programme, it is to be expected that the number of knowledge packages deposited in the GKH in the first year (2020) will be limited (between 2-8). The number of papers will be likely in the order of 50-100 and the number of shared scripts and software modules is expected to be between 10 and 20. Being on the conservative side, we thus set as performance measures for the first year of the GKH the following:

- Knowledge packages available : 4
- Papers deposited: 50.
- Shared scripts and software modules: 10.
- In-situ data sets associated with applications: 10.

7. Implementation Plan Acceptance Criteria

Upon viewing the proof-of-concept demonstration of the GKH involving the Sen2-Agri tool for deriving a crop mask, the GEO-XVI Plenary approved the proposal put forth by the GEO Executive Committee, requesting that the GEO Secretariat develop an implementation plan for further development of the GKH. The implementation plan was to be elaborated in consultation with the GEOSS Infrastructure Development Task Team and the GEO Programme Board, before being submitted to the 51st meeting of the Executive Committee for decision.

Therefore, acceptance criteria for this GKH implementation plan include:

- Initial review by GEOSS Infrastructure Task Team (mid-January 2020)
- Review by Programme Board (early February 2020)
- Final review by GEOSS Infrastructure Task Team (mid-February 2020)
- Submission to 51st Executive Committee meeting for decision (19-20 March 2020)

Each of these review phases will confirm that tasks and deliverables outlined in the implementation plan are acceptable, and whether there are further required technical processes, methods, tools, or performance benchmarks to be included in the plan prior to submission to the Executive Committee for final approval.

8. Glossary

AWS	Amazon Web Services
CC	Creative Commons
CCS	Cascade Style Sheets
CERN	European Organization for Nuclear Research
CONOPS	GEO Secretariat Concept of Operations
COPUOS	UN Committee on Peaceful Uses of Outer Space
DOI	Digital Object Identifier
EO	Earth Observations
FT	Foundational Task
GEE	Google Earth Engine

GEO	Group on Earth Observations
GEOSS	Global Earth Observation System of Systems
GKH	GEO Knowledge Hub
GPL	Gnu Public License
GWP	GEO Work Programme
HTML	Hypertext Markup Language
IP	Internet Protocol
IPR	Intellectual Property Rights
MIT	Massachusetts Institute of Technology
OGC	Open Geospatial Consortium
WMO	World Meteorological Organization

9. Documentation

In addition to documents cited elsewhere in this implementation plan, documents relevant to the development and delivery of the GKH include:

- GEO Strategic Plan 2016-2025: GEOSS Implementation
  http://earthobservations.org/documents/GEO_Strategic_Plan_2016_2025_Implementing_GEOSS.pdf

- GEO Rules of Procedure (March 2019 update)
  http://earthobservations.org/documents/GEO_Rules_of_Procedure.pdf

- 2020-2022 GEO Work Programme
  http://earthobservations.org/documents/geo16/GEO-XVI-7.2_2020-2022%20GEO%20Work%20Programme.pdf

- Results-Oriented GEOSS: A framework for transforming Earth observation data to knowledge for decision making
  http://earthobservations.org/documents/geo16/GEO-XVI-6.1(Rev1)_Results-Oriented%20GEOSS_A%20framework%20for%20transforming%20Earth%20observation%20data%20to%20knowledge%20for%20decision%20making.pdf

- Proposed Design and Proof of Concept of the GEO Knowledge Hub
  http://earthobservations.org/documents/geo16/GEO-XVI-6.2(Rev1)_Proposed%20Design%20and%20Proof%20of%20Concept%20of%20the%20GEO%20Knowledge%20Hub.pdf

- GEO Secretariat Concept of Operations (CONOPS)
  http://www.earthobservations.org/documents/excom/ec48/ExCom48-06_rev1%20GEO%20Secretariat%20Concept%20of%20Operations.pdf